Designing a Graph-based Multiple-level Association Mechanism to Discover the Learners' Preferences in Course Categories

Ko-Kang Chu
kirk@ms2.hinet.net
Dept. of Electronic Engineering, Chung-Yuan Christian Univ.
#22, Pu-Jen, Pu-Chung Li, Chung-Li, 320, Taiwan

and

Maiga Chang
maiga@ms2.hinet.net
National Science & Technology Program for eLearning in Taiwan, National Central Univ.
#300, Jung-Da Rd., Chung-Li, 320, Taiwan

and
Yen-Teh Hsia
hsia@ice.cycu.edu.tw
Dept. of Information and Computer Engineering, Chung-Yuan Christian Univ.
#22, Pu-Jen, Pu-Chung Li, Chung-Li, 320, Taiwan

ABSTRACT

This paper is trying to propose a mechanism that only according to the learner's course selection records (current and even past) to analyze the preference model of learner. Learner's interesting order of course categories should be able to find out based on what courses she/he had ever been took. In this paper, we use directed graph to analyze and generate the learner's interesting order of course categories. The preference model we developed with directed graph will be able to growth, merge and adapt to learners' interest changes. Finally, actual results come from two samples of undergraduate students in the department of Information and Computer Engineering, Chung-Yuan Christian Univ., are used for evaluating our theory.

Keywords: Association Rule, Data Mining, Graph Theory, Preference Model, eLearning

1. INTRODUCTION

The course selection records of learners can be thought as the course preferences of learners, although there are many issues influence the learners' choices such as "does this course easy to pass?", "when is the class?", "does the teacher a good teacher?" *etc.*. Since learners' choices are recorded by either course selection system or formal documents, it should be able to analyze via Data Mining techniques very easily. Data Mining techniques are currently quite common used for analyzing structured data in Business[9][11][13], Education[8][10] and AI fields.[1][7]

The most simple data mining algorithm is proposed by Agrawal and Srikant in 1994, which called Apriori.[2] Based on Apriori algorithm, the association rules, "A → B" for example, can be found from large databases. However, the scenario of selecting courses in school as Figure 1 shown below, each course has several characteristics which cause the course belongs to one course category and even several categories at the same time..

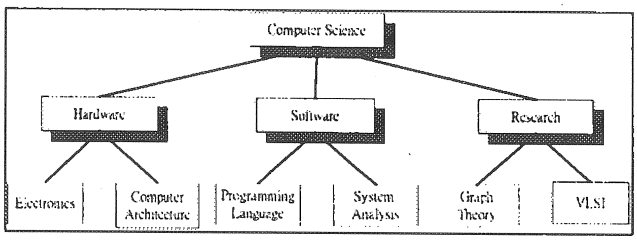


Figure 1. Category Hierarchy for Courses

Similar to our objective, other advanced algorithms are separately proposed by Han and Srikant in 1995, called multiple-level association rule.[6][12] However, from Example 1 below we know that not every course belongs to one category, most of them will involve more than one category. Therefore, a new methodology is needed to achieve our goal consequently.

Example 1. Given 6 courses and its related categories
Fuzzy: AI (95%), Research (85%), Math (70%)
Neural Net: AI (90%), Research (85%), Math (70%)
Computer Mathematics: Research (90%), Math (90%)
Graph Theory: Research (100%), Math (80%)
Statistics: Math (100%)
Data Mining: AI (95%), Research (90%)
There are several students who take different courses as following:
Ken: Fuzzy and Neural Networks
Janie: Fuzzy and Statistics
John: Data Mining and Neural Networks. □

As example described above, it will be not easy to know which category that Ken really likes and perhaps the category weights for each course are totally different from Ken's view.

Section 2 uses directed graph to model the learner's interesting order of course categories based on their selection records. An algorithm is developed and a simple accompany example is also provided in Section 3. Two samples of undergraduate students are invited to use our experiment system (a course selection system) built on Web and a questionnaire sheet is used to verify the preference models generated by our mechanism. The comparisons are made in Section 4 and what can be done for future researches will be discussed by Section 5.

2. INTERESTING ORDER OF COURSE CATEGORIES

As known each course will not just belong to only one course category, for example, course Fuzzy in Example 1 belongs to Artificial Intelligence, Research and even Mathematics. More than that, as Example 1 represented above the degrees of the relations among course and categories are quite variedly. Therefore, there are several works should be done before analyzing the interesting order of course categories which existed in the learner's mind:

i. defining course categories
ii. deciding what categories the course belongs to
iii. weighting the relations of course and categories

These three requirements described above could be achieved simply by inviting professionals' participation.

To complete the analysis of interesting order of course categories and find the preference model of learners, some notations are needed to be defined first. In our previous study, the relation matrix had been designed for representing both relations and degrees between courses and categories.[4][5] Example 2 takes Example 1 and makes a simple introduction to the pre-defined notations below.

Example 2. (based on the Example 1) Given 6 courses and its related categories

Courses,

$C = \{Fuzzy, NN, CMath, GTheory, Statistics, DM\}$

Categories, $T = \{t_1, t_2, t_3\} = \{AI, Research, Math\}$

$$CM = \begin{bmatrix} f_{AI}(Fuzzy) & f_{Research}(Fuzzy) & f_{Math}(Fuzzy) \\ f_{AI}(NN) & f_{Research}(NN) & f_{Math}(NN) \\ f_{AI}(CMath) & f_{Research}(CMath) & f_{Math}(CMath) \\ f_{AI}(GTheory) & f_{Research}(GTheory) & f_{Math}(GTheory) \\ f_{AI}(Statistics) & f_{Research}(Statistics) & f_{Math}(Statistics) \\ f_{AI}(DM) & f_{Research}(DM) & f_{Math}(DM) \end{bmatrix}$$

$$CM = \begin{bmatrix} 0.95 & 0.85 & 0.70 \\ 0.90 & 0.85 & 0.70 \\ 0 & 0.90 & 0.90 \\ 0 & 1 & 0.80 \\ 0 & 0 & 1 \\ 0.95 & 0.90 & 0 \end{bmatrix}$$ □

After introducing the relative relatedness function f_i and the category matrix CM, the interesting order of course categories, λ_{course}, can be constructed. Here, we use two symbols to indicate different orders: ">" means more important and "~" means undecided (still can not decide at present time).

Example 3. (according to Example 2) The learners' interesting order could be told when she/he takes

i. the course Neural Networks (NN for short):

$$f_{AI}(NN) = 0.9 > f_{Research}(NN) = 0.85 > f_{Math}(NN) = 0.70$$
$$\Rightarrow \lambda_{NN} = AI > Research > Math$$

ii. the course Computer Mathematics (CMath for short):

$$f_{Research}(CMath) = 0.90 \sim f_{Math}(CMath) = 0.90$$
$$\Rightarrow \lambda_{CMath} = Research \sim Math \,.\, \square$$

As Example 3 demonstrated above, after one course is taken by learner we probably could say she/he has different degree of interest in different course categories. However, it is obviously that several courses would be taken by one learner at the same semester. Under such kind of situations, how can we deduct the preference model of the learner with so many various interesting orders from different courses she/he chooses?

As mentioned above, not only data mining techniques, but also the directed graphs are needed to integrate the interesting orders while modeling the learner's preferences. Figure 2 indicates how to use graph to represent the interesting order. The Circles in Figure 2 represent categories and the arrows indicate important degree.

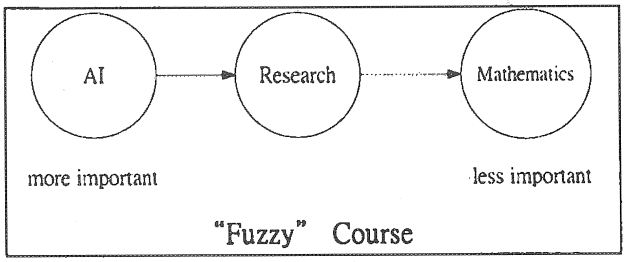


Figure 2. The interesting order for the course of Fuzzy

Beside ">" and "~" shown in Example 3, there is one more symbol – implicitly more important ($\overline{>}$), is also needed to introduce when we try to merge two interesting orders. Let's take Example 4 as an example.

Example 4. If two courses α and β are taken by a learner, then two interesting orders λ_α and λ_β could be generated as

$$\lambda_\alpha = AI$$
$$\lambda_\beta = AI \sim Research$$
$$\Rightarrow \lambda_\beta = AI \,\overline{>}\, Research$$

Therefore, the interesting order in learner's mind shall be "category AI is implicitly more important than category Research." □

3. ALGORITHM OF MODELING LEARNER'S PREFERENCES

At the beginning of this Section, we first define several graphs for representing the different interesting orders in order to design the algorithm of modeling the learner's preference. In Figure 3, an interesting order of course categories, λ, is mapping to a directed graph, $G_d(\lambda)$. Once the graph representations for the interesting order are defined, a five steps algorithm is designed to construct the preference models of learners. A complete example, Example 5, will be given as a demonstration for the algorithm.

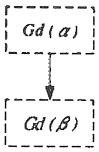
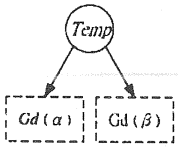
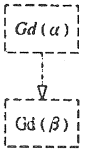
$\lambda =$	$G_d(\lambda)$	Description
t_m		If there is only one category t_m belonged to λ, then this type of node is called *SINGLE*.
$\lambda_\alpha > \lambda_\beta$		The solid arrow depicts the relationship between two nodes is "more important."
$\lambda_\alpha \sim \lambda_\beta$		If $\lambda = \lambda_\alpha \sim \lambda_\beta$, an extra *Temp* node then is used to link both λ_α and λ_β; if there is a λ_γ, which $\lambda_\gamma > \lambda_\alpha$ and $\lambda_\gamma > \lambda_\beta$, then the *Temp* node can be replaced by the $G_d(\lambda_\alpha)$ node.
$\lambda_\alpha \bar{>} \lambda_\beta$		The hollow arrow depicts the relationship between two nodes is "implicitly more important"; this type of edge is called "implicit edge."

Figure 3. Graph expressions of the interesting orders

Example 5. Given six interesting orders

$\lambda_1 = t_2$ $\quad\quad$ $\lambda_4 = t_4$

$\lambda_2 = t_2 > t_3 > t_1$ $\quad\quad$ $\lambda_5 = t_5 > t_1 \sim t_3$

$\lambda_3 = t_4 \sim t_5$ and $\lambda_6 = t_6$

what would be the preference model (merged interesting order) of the learner? □

Step 1. Drawing the corresponding graphs, $G_d(\lambda_i)$

Let $D = \bigcup_{\forall i} G_d(\lambda_i)$, let $V = \{v_1, v_2, \ldots, v_m\}$ be the vertex set of D, and v_i is an element in course category set T. As Figure 4(a) shows below.

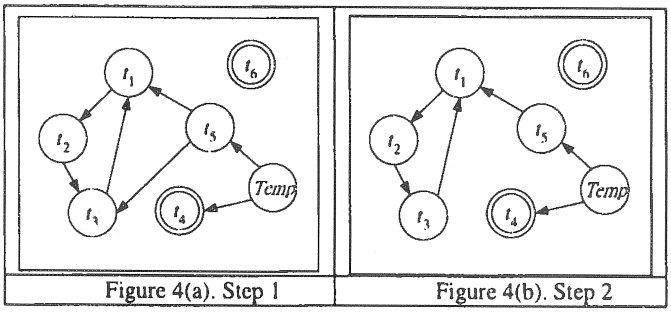
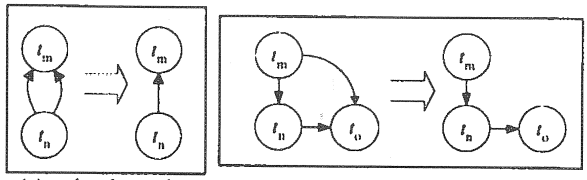


Figure 4(a). Step 1 $\quad$ Figure 4(b). Step 2

Step 2. Deleting unnecessary edges

If there is more than one edge between any two different vertices, $\overline{v_i v_j}$, then delete it. Besides, the transitivity relation should be also considered. Figure 5 summaries these two conditions.



(a) redundant edge $\quad$ (b) transitive edge

Figure 5. Unnecessary edges

Step 3. Grouping those vertices form cycles

Taking those vertices that will cause cycles and grouping them with a single vertex, as Figure 6(a) shown below.

Step 4. Adding implicit edges

i. $V_S = \{v_i \mid v_i \in SINGLE\}$

ii. Mapping D to an undirected graph U (every directed edges change to undirected one)

iii. For any vertex $v_i \in V_S$ if there exists a path from v_i to v_j and $\overline{v_i v_j} \notin Edge(U)$, then add a directed edge $\overline{v_i v_j}$ to D. Figure 6(b) demonstrates such operations described above.

iv. Repeat Step 2, deleting unnecessary edges as Figure 7(a) shown.

Step 5. Deleting *Temp* vertex and constructing order

i. Deleting the *Temp* vertices exist in D.

ii. Constructing preference order by going through those directed edges in D.

iii. Unfolding those grouping vertices combined by Step 3, and the preference relations among those vertices are "undecided."

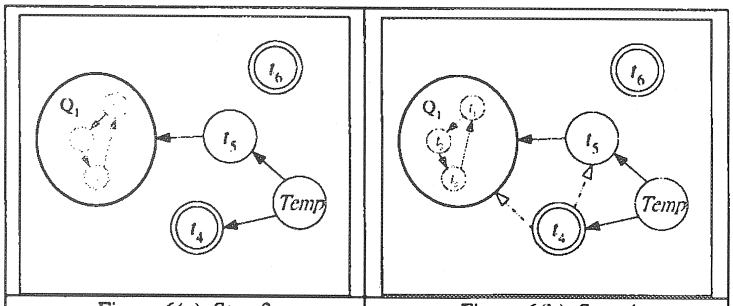


Figure 6(a). Step 3 $\quad$ Figure 6(b). Step 4

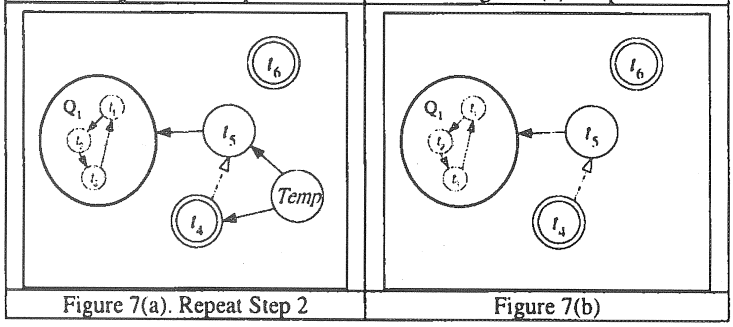


Figure 7(a). Repeat Step 2 $\quad$ Figure 7(b)

After applying the algorithm designed above, the final preference model of the learner in Example 5 could be found as $\lambda = t_6 \sim (t_4 \bar{>} t_5 > Q_1) = t_6 \sim (t_4 \bar{>} t_5 > t_1 \sim t_2 \sim t_3)$.

4. EXPERIMENTS AND RESULTS

An experiment system is built on the Web to let learners choose courses online.[3] The course recommendation system uses the algorithm proposed in this paper to make course suggestions for learners after they took courses. Figure 8 describes the operational process of the course recommendation system and Figure 9 represents several snapshots of the system. Figure 10 and Figure 11 represent two important steps in our system.

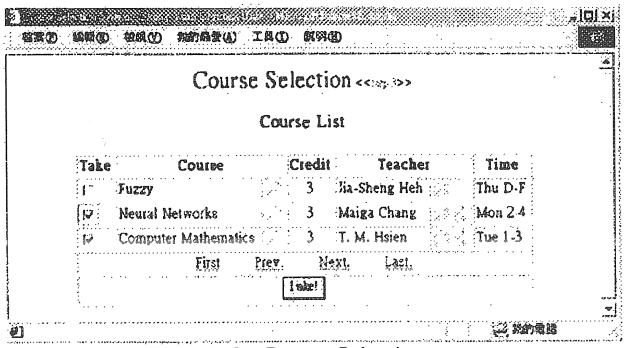


Figure 10. Course Selection step

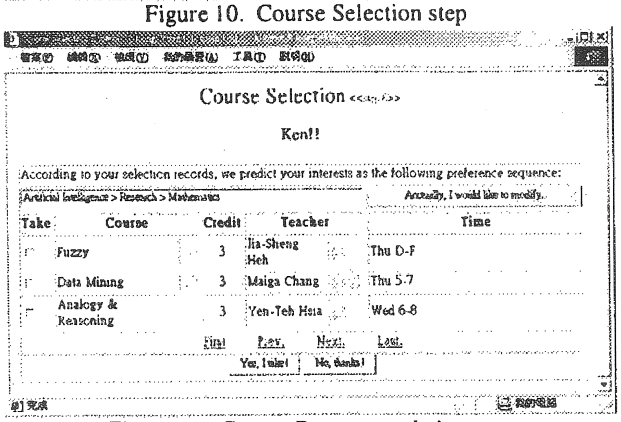


Figure 11. Course Recommendation step

Via the course selection system we collected course selection records of 4 consecutive school terms from the senior college students in the department of Information and Computer Engineering, and also conducted a questionnaire for the same two classes of students (the class of 2001 and 2002). About the class of 2001 who have graduated already, although the system has analyzed the course selection records of 127 students, out of 83 questionnaires we sent out, only 34 have responded. About the class of 2002, the system has analyzed the course selection records of 102 students, and all 102 of them have responded to our questionnaire. The accuracy rate of the preference sequence between our deduction outcomes and students' questionnaire returns can be seen in Figure 12 and Figure 13.

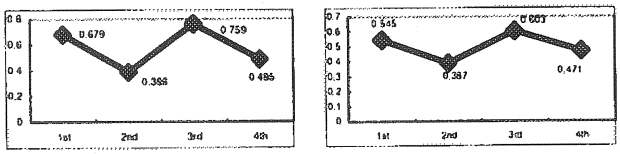
Figure 12 lists partial of preference sequences from the feedback provided by students. The data stored in the second column in Figure 12 is the actual preference orders of students that are gathered from questionnaires. Column 3 to 7 represent the preference sequences of each semester analyzed by using our mechanism automatically during their 2nd year (2nd semester) to 4th year (2nd semester) in college. (We have 2 semesters in each school year)

ID	Questionnaires Results	2nd Year 2nd semester	3rd Year 1st semester	3rd Year 2nd semester	4th Year 1st semester	4th Year 2nd semester
101	B>A>F>C>D>E	F>D~E		D		E>A~B
103	F>E>A>D>B>C		B>A>C	D		
105	F>E>A>D>B>C		B~E	D		
107	F>E>D		E~(B>A>C)	D		B>E
120	F>E>C>B>A>D	F>D>E		D		E>A>B
123	F>E>A>D>B>C	E	B>A>C	D		A~B~E
126	E>D>F>B>C>A					C~(B>E)
129	F>E>D>A>C>B		B>A>C	D		
140	F>E>A>D>B>C		E	D		B>E
152	F>E>B>A>C>D	F>D>E				D>B>E
154	F>E>A>C>D>B		B	D~(B>E)		
155	E>D>F>B>C>A	F>E>D				
158	F>E>D>A~B~C		B>A>C	D		
204	F>D>E>B>C>A		F>D~(B>A>C)			A>B>E
207	C>B>A>F>E>D	E			E>B	D~(B>E)
208	C>B>A>F>E>D	E		A>B	E>B	D~(B>E)
210	C>B>A>F>E>D		E~(B>A>C)	A>B	E>B	
211	F>E>D>B>A>C		E~(F>A~B~D)	B>C>A		
ID	Questionnaires Results	2nd Year 2nd semester	3rd Year 1st semester	3rd Year 2nd semester	4th Year 1st semester	4th Year 2nd semester
214	E>F>D>B>A>C					D~(B>E)
217	F>E>D>B>A>C			D		B>E
218	F>E>D>C>B>A	E			B~E	C
223	A>C>F>E>B>D					A~B~C
225	A>C>F>E>B>D		F>D~(B>A>C)	D		D>B
226	F>E>D>B>A>C					D~(A>B)
228	F>E>A>B>D>C				E	
232	C>B>A>F>E>D		B>A>C			D~(B>E)
233	F>E>D>C>B>A		B>A>C			
235	F>E>A>B>D>C	E		D		
237	F>A>B>E>D>C		E~F>A~B~D			E>B
238	F>A>B>E>D>C		B			
239	F>E>A>B>D>C	F>D~E	F>D>E>B	D		
240	F>E>D>C>B>A		F>D>E	A>B		
244	F>E>A>B>D>C		F>D~(B>A>C)			
246	F>A>B>E>D>C	E			E>B	B>E~(C>A)

Figure 12. Preference sequence of predicted by our system and feedback from students

Cells with **black** background in Figure 12 mean the interesting order of students match our analysis results exactly. Cells with **gray** background indicate there should be a little difference between the interesting order in students' minds and our analysis results. Cells with **white** background mean our predictions are somehow different to the interesting order in students' minds.

Although the accuracy rates of the 4th year of college in both Figure 13(a) and 13(b) are not good enough, a hypothesis is made and can be verified with Figure 14. The hypothesis we made here is "some of students are applying for or already being accepted by graduate schools, they tend to follow graduate school instructor's guidance, which sometimes deviate considerably from their prior perception of the order of relative importance, to take courses more related to their future studies."



(a) class 2001 (b) class 2002

Figure 13. Accuracy rate calculated from Figure 10

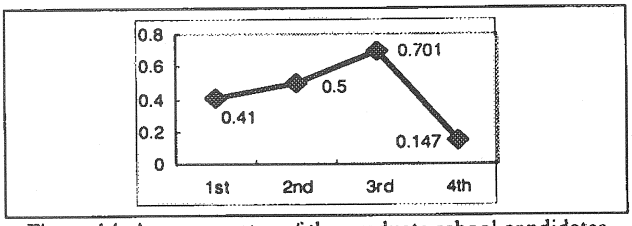


Figure 14. Accuracy rates of the graduate school candidates among class 2002

5. CONCLUSIONS

This paper proposes a graph-based multiple-level association algorithm for modeling the learners' interests in course categories, with this algorithm an experiment system is built to evaluate our theory. From now on, there are still several works we shall do:

i. How to find course categories classified by students?

ii. Time series analysis should be made, which means can we predict the students' change of interest from year to year.

iii. Course development: can we develop or plan a series of courses depends on the student's major interests? (a sort of professional education.)

6. REFERENCES

[1] Agrawal, R., Imielinski, T., & Sqami, A. (1993). Mining Association Rules between Sets of Items in Large Databases. **Int'l Conf. Management of Data (SIGMOD'93)**, ACM, Washington, DC. 207-216.

[2] Agrawal, R., & Srikant, R. (1994). Fast Algorithms for Mining Association Rules. **Int'l Conf. Very Large Databases (VLDB'94)**, Santiago, Chile. 487-499.

[3] Chu, Ko-Kang, Chang, Maiga, & Hsia, Yen-Teh. (2003) Designing a Course Recommendation System on Web based on the Students' Course Selection Records, **World Conference on Educational Multimedia, Hypermedia and Telecommunications (ED-Media 2003)**, Honolulu, Hawaii, USA, June 23-28, 2003. 14-21

[4] Chu, Ko-Kang, & Hsia, Yen-Teh. (2001). Formula of Relative Importance for Modeling User Behavior. **Proceedings of the Sixth Conference on Artificial Intelligence and Applications (TAAI 2001)**, Nov. 9, 2001, Kaohsiung, Taiwan. 444-449.

[5] Chu, Ko-Kang, & Hsia, Yen-Teh. (2002). Deducing the Importance of Course Categories from Course Selection Records. **Chung-Yuan Journal**, 30 (3), 401-410.

[6] Han, J., & Fu, Y. (1995). Discovery of Multiple-Level Association Rules from Large Data Bases, **Int'l Conf. Very Large Databases (VLDB'95)**, Zurich, Switzerland. 420-431.

[7] Han, J., & Kamber, M. (2001). **Data Mining Concepts and Techniques**. Morgan Kaufmann.

[8] Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H., & Verkamo, A. I. (1994). Finding Interesting Rules from Large Sets of Discovered Association Rules. **Int'l Conf. Information and Knowledge Management**, 1994, Gaithersburg, MD. 401-408.

[9] Müller, R. J., & Yang, Y. (1997). Association Rules over Interval Data. **Int'l Conf. Management of Data (SIGMOD'97)**, ACM, Tucson, AZ. 452-461.

[10] Mannila, H., Toivonen, H., & Verkamo, A. I. (1994). Efficient Algorithms for Discovering Association Rules. **AAAI'94 Workshop Knowledge Discovery in Database (KDD'94)**, *AAAI*, Seattle, Washington. 181-192.

[11] Ozden, B., Ramaswamy, S., & Silberschatz, A. (1998). Cyclic Association Rules. **Int'l Conf. Data Engineering (ICDE'98)**, Orlando, FL. 412-421.

[12] Srikant, R., & Agrawal, R. (1995). Mining Generalized Association Rules. **Int'l Conf. Very Large Data Bases (VLDB'95)**, Zurich, Switzerland. 407-419.

[13] Srtkant, R., & Agrawal, R. (1996). Mining Quantitative Association Rules in Large Relational. **Int'l Conf. Management of Data (SIGMOD'96)**, Montreal, Canada. 1-12.

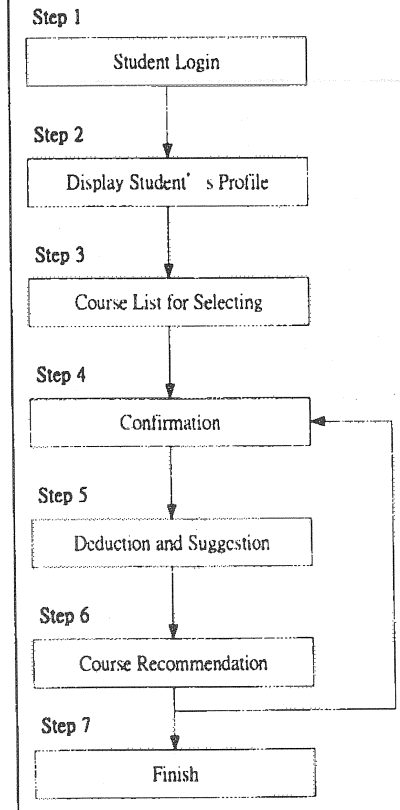


Figure 8. Process of Course Recommendation System

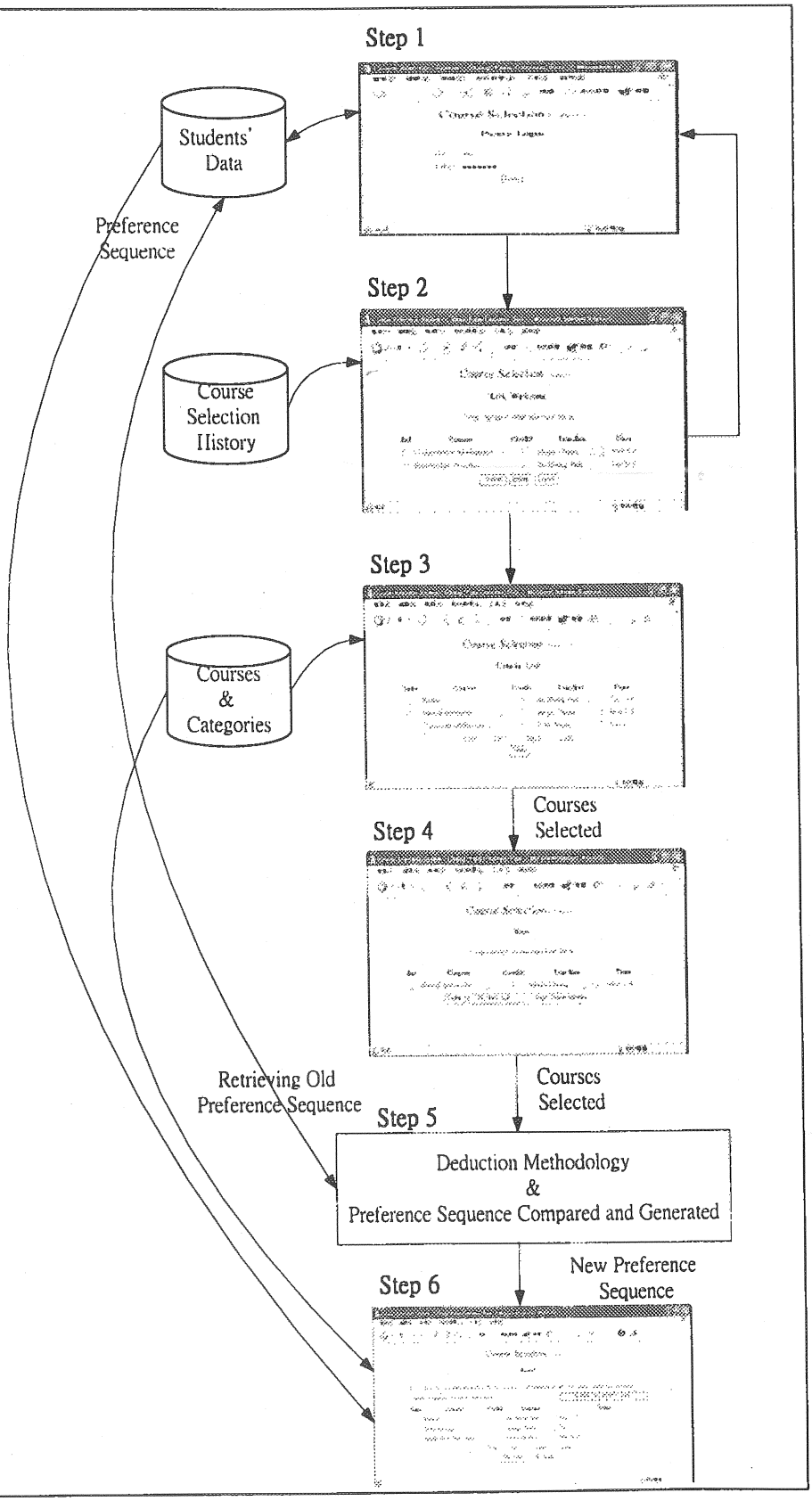


Figure 9. Architecture of Course Recommendation System